PERGAMON

Topology 39 (2000) 773–787

www.elsevier.com/locate/top

A perfect Morse function on the moduli space of flat connections☆

Michael Thaddeus*

Department of Mathematics, Columbia University, 2990 Broadway, New York, NY 10027, USA

Received 21 December 1998; accepted 22 January 1999

Abstract

The moduli space of flat SU(2) connections on a punctured surface, having prescribed holonomy around the punctures, is a compact smooth manifold if the holonomy is generic. This paper gives a direct, elementary proof that the trace of the holonomy around a certain loop determines a Bott–Morse function on the moduli space which is perfect, meaning that the Morse inequalities are equalities. This leads to an attractive recursion for the Betti numbers of the moduli space, which agrees with the Harder–Narasimhan formula in the case of one puncture with holonomy -1.

MSC: 58E05

Keywords: Flat connections; Riemann surfaces; Morse theory; Symplectic geometry

Let X be a compact surface of genus g, and let M_g be the moduli space of flat SU(2) connections on X having holonomy $-I$ around a single puncture p. Let $a_1, b_1, a_2, b_2, \dots, a_g, b_g$ be the usual generators for $\pi_1(X \backslash p)$, and define a real-valued function on M_g by assigning to a flat connection the trace of its holonomy around a_g. This paper will give an elementary, direct proof that this function is a *perfect* Bott–Morse function, that is, one whose Morse inequalities are equalities. This leads to a new derivation of the well-known Harder–Narasimhan formula for the Betti numbers of the moduli space.

* Corresponding author. Tel.: + 212-854-4308; fax: + 212-854-8962.
E-mail address: thaddeus@math.columbia.edu (M. Thaddeus)
☆ Partially supported by NSF grants DMS-9500964 and DMS-9808529.

PII: S0040-9383(99)00032-4

The Harder–Narasimhan formula has previously been calculated by several methods. Most of them involved the deep theorem of Narasimhan and Seshadri [22] identifying flat connections with stable holomorphic bundles. For example, Harder and Narasimhan [13] applied the Weil conjectures, counting algebraic vector bundles on curves over finite fields to deduce the result. Another approach, due to Atiyah and Bott [1], used two-dimensional Yang–Mills theory to construct M_g as the symplectic quotient of an infinite-dimensional affine space by the gauge group.

More recently, the Bott–Morse function discussed here was studied by Jeffrey and Weitsman [17]. It is straightforward to determine the critical points of this function (cf. 1.5), and hence to compute one side of the Morse inequalities. Jeffrey and Weitsman noted that the expression thus obtained equals the Harder–Narasimhan formula. If the latter is regarded as known, this constitutes an a posteriori proof that the function is perfect.

This paper will give an a priori proof that the function is perfect, using only relatively elementary facts from finite-dimensional symplectic geometry. It therefore provides a down-to-earth proof of the Harder–Narasimhan formula, avoiding high technology such as the Weil conjectures or Yang–Mills theory. It is much closer in spirit to the original computation of the Betti numbers, due to Newstead [12,23].

A few years ago, Mark Hoyle pointed out to the author that the a posteriori argument remains valid for the space of flat SU(2) connections whose holonomy around p lies in any fixed conjugacy class (except that of I when there are reducibles). In this case the Betti numbers are easily determined from the Harder–Narasimhan formula. The present a priori argument extends without change to this space. Even better, it applies more generally to the space of flat SU(2) connections on X with any finite number of punctures, provided only that the conjugacy classes of the holonomies around the punctures are fixed so that there are no reducibles. Meanwhile Hoyle has obtained formulas for the Betti numbers in this case too and shown that the a posteriori argument extends [14]. The results of this paper therefore provide an independent proof of his formulas.

A natural question for further investigation is how much can be said for other Lie groups than SU(2). One could replace the trace with any real-valued smooth class function. The proof of the main theorem does not carry over in full, but it might imply some partial results. Furthermore, the a posteriori question should be tractable.

The organization of the paper is simple. Section 1 reviews what little background material is needed: the construction of the moduli space as a space of representations of $\pi_1(X)$, and some results of Goldman and Frankel. The main theorem is stated, and the Poincaré polynomial of the moduli space is derived. Section 2 describes the spectral sequence, canonically associated to a Bott–Morse function $f: M \to \mathbb{R}$, which abuts to $H^*(M, \mathbb{Q})$. Section 3 is devoted to the proof of the main theorem. The basic strategy is to prove that all differentials vanish in the spectral sequence by combining the perfection of moment maps with a parity argument. Section 4 explains how to generalize the main theorem to the case of a surface with many punctures, recovering the formulas of Hoyle. Finally, Section 5 explains why the methods of Sections 2 and 3 immediately show that the integral cohomology is torsion-free, and yield information about the U(2) case as well.

Unless otherwise specified, all cohomology is with rational coefficients.

This work was announced in lectures given in Odense and Barcelona [25]. I wish to apologize for the delay in publication.

1. The moduli space of flat SU(2) connections

The moduli space M_g which interests us can be defined simply as follows. Let $\mu_g : \mathrm{SU}(2)^{2g} \to \mathrm{SU}(2)$ be given by $(A_1, B_1, \dots, A_g, B_g) \mapsto -\prod_i [A_i, B_i]$. Then I is a regular value of μ_g. (This was first shown by Igusa [15]; it is also a special case of Proposition 4.1 below.) Hence $\mu_g^{-1}(I)$ is smooth; moreover $\mathrm{SU}(2)/\pm I$ acts freely on it by conjugation, so $M_g = \mu_g^{-1}(I)/\mathrm{SU}(2)$ is a smooth compact $(6g-6)$-dimensional manifold.

At any point $\rho \in M_g$, the tangent space is naturally isomorphic to the first cohomology of the complex

$$\mathfrak{su}(2) \to \mathfrak{su}(2)^{2g} \to \mathfrak{su}(2),$$

where the former map is the derivative of conjugation at (A_i, B_i), and the latter is the derivative of μ_g. Choose any representative for ρ in $\mu_g^{-1}(I)$ and, by abuse of notation, denote it again by ρ. Then $\mathrm{ad}\,\rho$ defines a representation of $\pi_1(X)$ on $\mathfrak{su}(2)$, and the cohomology mentioned above is none other than the group cohomology $H^1(\pi_1(X), \mathrm{ad}\,\rho)$.

Since X is an Eilenberg–Mac Lane space, $H^2(\pi_1(X), \mathbb{R}) = H^2(X, \mathbb{R}) = \mathbb{R}$. Combining the cup product with the symmetric form $\langle a, b \rangle = -\frac{1}{2}\mathrm{tr}\, ab$ gives a nondegenerate antisymmetric map

$$H^1(\mathrm{ad}\,\rho) \otimes H^1(\mathrm{ad}\,\rho) \xrightarrow{\cup} H^2(\mathrm{ad}\,\rho \otimes \mathrm{ad}\,\rho) \xrightarrow{\langle\,,\,\rangle} H^2(\mathbb{R}) = \mathbb{R},$$

which determines a nondegenerate 2-form on M_g.

Theorem 1.1 (Goldman). *This 2-form is closed.*

Thus, M_g becomes a symplectic manifold. Goldman's original proof of this theorem [8] used infinite-dimensional quotients in the style of Atiyah and Bott. Since then, purely algebraic, finite-dimensional proofs have been provided by Karshon [18] and Weinstein [26]; see also Guruprasad et al. [11].

Now let $f : M_g \to [-1, 1]$ be given by $(A_i, B_i) \mapsto \frac{1}{2}\mathrm{tr}\, A_g$, which is well-defined since the trace is conjugation-invariant. Then U(1) acts on $f^{-1}(-1, 1)$ as follows. If $A_g \neq \pm I$, then there is a unique homomorphism $\phi : \mathrm{U}(1) \to \mathrm{SU}(2)$ such that $A_g \in \phi(\{\mathrm{Im}\, z > 0\})$. Let $\lambda \cdot (A_i, B_i) = (A_1, B_1, \dots, A_g, B_g \cdot \phi(\lambda))$.

Proposition 1.2 (Goldman). *This action preserves the symplectic form, and it has moment map* $-i \arccos f$.

See Kirwan [19] for the definition of a moment map, and Goldman [9] and Jeffrey–Weitsman [16] for a proof.

If the U(1)-action were global on M_g, the following result [7] would immediately imply that $\arccos f$ was perfect.

Theorem 1.3 (Frankel). *The moment map of a* U(1)-*action on a compact symplectic manifold* (*times* i) *is a perfect Bott–Morse function.*

In the present case, however, the U(1)-action on $f^{-1}(-1,1)$ does not extend over $f^{-1}(\pm 1)$, and $\arccos f$ is not even differentiable there. Nevertheless, the following theorem is true, and will be proved in Section 3.

1.4. Main Theorem. *The map f is a perfect Bott–Morse function on M_g.*

To see what this means concretely, let us identify the critical submanifolds and their indices, assuming for the moment that f is a Bott–Morse function.

1.5. The critical submanifolds of f. First, let $S_1 = f^{-1}(-1)$. As the absolute minimum of f, S_1 is of course a critical submanifold. It is exactly the locus where $A_g = -I$; hence B_g may be arbitrary, and the product of the first $g-1$ commutators must be $-I$, so $S_1 = (\mu_{g-1}^{-1}(I) \times \mathrm{SU}(2))/\mathrm{SU}(2)$. The natural projection $\pi : S_1 \to M_{g-1}$ makes S_1 into an SU(2)-bundle over M_{g-1}. This is an adjoint, not a principal bundle, so it may have a section without being trivial. Indeed, $B_g = I$ determines such a section; hence the Euler class vanishes and so by the Gysin sequence $H^*(S_1) = V \oplus \tau V$, where $V = \pi^* H^*(M_{g-1})$ and $\tau \in H^3(S_1)$ is the Poincaré dual of the locus where $B_g = I$. As the absolute minimum, S_1 of course has index 0.

Exactly the same is true of $S_3 = f^{-1}(1)$, except that it is the absolute maximum of f, the locus where $A_g = I$, and hence has index equal to its codimension, which is 3.

Within $f^{-1}(-1,1)$, on the other hand, the critical points of f coincide with those of $\arccos f$, the moment map for the U(1)-action. They are therefore exactly the fixed points of that action, and hence are represented by $2g$-tuples $(A_i, B_i) \in \mathrm{SU}(2)^{2g}$ that are conjugate to $(A_1, B_1, \ldots, A_g, B_g \cdot \phi(\lambda))$ for all $\lambda \in \mathrm{U}(1)$. It is straightforward to check that these are all conjugate to $2g$-tuples such that

$$A_g = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \qquad B_g = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

and the remaining A_i and B_i are diagonal. Hence the only other critical value is $\frac{1}{2}\mathrm{tr}\, A_g$, which is 0, and the corresponding critical set S_2 is a $2g-2$-torus. Because there is an involution $A_g \mapsto -A_g$ on M_g changing the sign of f, the index of S_2 must be half the rank of its normal bundle, or $2g-2$.

Incidentally, S_1 and S_3 are empty when $g = 1$, and everything is empty when $g = 0$, but these special cases will not affect our arguments.

The main theorem therefore implies the so-called Harder–Narasimhan formula.

Corollary 1.6. *The Poincaré polynomial of M_g is*

$$P_t(M_g) = \frac{(1+t^3)^{2g} - t^{2g}(1+t)^{2g}}{(1-t^2)(1-t^4)}.$$

Proof. Since M_0 is empty, $P_t(M_0) = 0$ as desired. It follows from the theorem and the discussion above that

$$\begin{aligned} P_t(M_g) &= P_t(S_1) + t^3 P_t(S_3) + t^{2g-2} P_t(S_2) \\ &= (1+t^3)^2 P_t(M_{g-1}) + t^{2g-2}(1+t)^{2g-2}. \end{aligned}$$

This gives a recursion for $P_t(M_g)$, which is satisfied by the formula. □

2. The spectral sequence of a Bott–Morse function

Let M be a compact manifold and $f: M \to \mathbb{R}$ a *Bott–Morse function*, that is, a smooth function whose critical set is a disjoint union of submanifolds on whose normal bundles the Hessian is nondegenerate. For f to be perfect is clearly equivalent to the vanishing of all differentials in the spectral sequence described below.

The set of critical values of f is finite, so the critical set of f is a disjoint union of submanifolds $S_1, S_2, \dots, S_n \subset M$ such that f is constant on each S_i and $f(S_i) < f(S_j)$ for $i < j$. Choose $x_j \in \mathbb{R}$ satisfying $x_0 < f(S_1) < x_1 < f(S_2) < x_2 < \cdots < f(S_n) < x_n$ and let $U_j = f^{-1}(x_0, x_j)$, so that M is filtered by open sets $\emptyset = U_0 \subset U_1 \subset \cdots \subset U_n = M$. The complex of singular chains on M is then filtered by the support of the chain:

$$0 = C_*(U_0) \subset C_*(U_1) \subset \cdots \subset C_*(U_n) = C_*(M).$$

Moreover, this filtration is preserved by the differential.

Taking duals yields a cofiltration on the group of cochains, which is also preserved by the differential. This is exactly the raw material needed to construct a spectral sequence, and the whole machine runs smoothly. Except that everything is dualized because of the cofiltration, it is just as described by Bott and Tu [6]. As in their equation 14.3, there is a short exact sequence

$$0 \to B \to A \xrightarrow{\pi} A \to 0$$

where $A = \oplus_j C^*(U_j)$ and $\pi: C^*(U_j) \to C^*(U_{j-1})$ is the restriction of cochains. This leads to an exact couple

$$\begin{tikzcd} \oplus H^*(U_j) \arrow[rr] & & \oplus H^*(U_j) \arrow[dl] \\ & \oplus H^*(U_j, U_{j-1}) \arrow[ul] & \end{tikzcd}$$

whose derived couples abut to $H^*(M)$. On the other hand, the Morse lemma implies that up to homotopy, U_j is a CW complex obtained from U_{j-1} by attaching, along its boundary, the disc bundle E_j associated to the *negative normal bundle* consisting of the negative definite subspaces of the Hessian of f on S_j. So by excision, $H^*(U_j, U_{j-1}) \cong H^*(E_j, \partial E_j)$.

On the other hand, suppose that S_j and E_j are orientable for each j, and choose orientations. Then the cup product with the Thom class induces the *Thom isomorphism* $H^*(E_j, \partial E_j) \cong H^*(S_j)$. The spectral sequence then can be regarded as a sequence of differentials $d_1, d_2, \dots$ where d_1: $\oplus H^*(S_j) \to \oplus H^*(S_j)$ and $d_{i+1}: H^*(d_i) \to H^*(d_i)$.

By the definition of the differentials of an exact couple, the first differential d_1 is the direct sum of maps $H^*(S_j) \to H^*(S_{j+1})$ induced by the upper part of the following diagram:

$$\begin{tikzcd}
H^*(S_j) \arrow[dr] & & & \\
& H^*(U_j, U_{j-1}) \arrow[d] & & \\
H^*(U_{j+1}) \arrow[r] \arrow[d] & H^*(U_j) \arrow[r] & H^*(U_{j+1}, U_j) \arrow[dr] & \\
H^*(U_{j+2}, U_{j+1}) \arrow[dr] & & & H^*(S_{j+1}) \\
& H^*(S_{j+2}) & &
\end{tikzcd} \tag{2.1}$$

Here all the maps are induced by inclusion, except the diagonal arrows, which are Thom isomorphisms.

An element of $H^*(S_j)$ is thus in the kernel of the first differential if and only if it maps to $0 \in H^*(U_{j+1}, U_j)$. By exactness of the middle row, it comes from some element of $H^*(U_{j+1})$, and its image in $H^*(S_{j+2})$ is the value of the second differential.

The higher differentials can now be described in the same manner, applying the argument of the previous paragraph inductively.

The contents of this section go back to Bott [5].

3. Proof of the main theorem

Let M_g and $f: M_g \to \mathbb{R}$ be as in Section 1. To prove the main theorem, we will show first that f is a Bott–Morse function, then that it is perfect. The former task is accomplished with three lemmas; the latter will occupy the remainder of the section.

Lemma 3.1. *For $g \geqslant 2$, the map $\mu_g^{-1}(I) \to \mathrm{SU}(2) \times \mathrm{SU}(2)$ given by $(A_i, B_i) \mapsto (A_g, B_g)$ is a submersion at the locus where $A_g = \pm I$.*

Proof. This means that the infinitesimal map $\ker D\mu_g \to \mathfrak{su}(2) \times \mathfrak{su}(2)$ induced by projection of $\mathrm{SU}(2)^{2g}$ on the last two factors is surjective when $A_g = \pm I$. Direct computation shows that in this case,

$$D\mu_g(a_1, \dots, b_g) = D\mu_{g-1}(a_1, \dots, b_{g-1}) + D\mu_1(a_g, b_g).$$

But as mentioned before, I is a regular value of μ_{g-1}. Hence $D\mu_{g-1}$ is surjective, so there exist $(a_i, b_i) \in \ker D\mu_g$ having any desired values for a_g and b_g. $\square$

Lemma 3.2. *There is an* $\mathrm{SU}(2)$*-equivariant diffeomorphism between a neighborhood of $S_1 \subset \mu_g^{-1}(I)$ and a neighborhood of*

$$\mu_{g-1}^{-1}(I) \times \{-I\} \times \mathrm{SU}(2) \subset \mu_{g-1}^{-1}(I) \times \mathrm{SU}(2) \times \mathrm{SU}(2),$$

identifying the map of the previous lemma with the projection on the last two factors. Likewise for S_3 if $\{I\}$ is substituted for $\{-I\}$.

Proof. This follows immediately from the previous lemma using the equivariant version of the tubular neighborhood theorem for submanifolds [2, I Theorem 2.1.1] and the inverse function theorem. $\square$

Lemma 3.3. *The function f is a Bott–Morse function.*

Proof. By definition, we need to show that the Hessian on the normal bundle to every critical submanifold is nondegenerate. By Frankel's result, Theorem 1.3, $\arccos f$ is a Bott–Morse function on $f^{-1}(-1,1)$, and hence so is f. Therefore the Hessian of S_2 is nondegenerate. In a neighborhood of S_1, f is locally the composite with the diffeomorphism from Lemma 3.2 of the trace map on the first SU(2) factor. But if SU(2) is identified with the 3-sphere of unit quaternions, then the trace map is simply a linear projection, whose Hessian at $\pm I$ is certainly nondegenerate. Hence the Hessian of f is nondegenerate on S_1. The case of S_3 is similar. $\square$

3.4. The symplectic cut $\hat{M}_g$. It remains to show that f is perfect. As a first step towards this goal, we construct a slight modification of M_g on which the U(1)-action becomes global and hence the moment map is perfect. Take the 2-sphere S^2 with the standard U(1)-action given by rotation, and normalize the moment map so that its image is $[-\frac{2}{3},\frac{2}{3}]$. Then let $\hat{M}_g$ be the symplectic quotient of $f^{-1}(-1,1)\times S^2$ by the diagonal U(1)-action. This is a simple example of a *symplectic cut* of M_g, as introduced by Lerman [20]. As a set, it is just $f^{-1}[-\frac{2}{3},\frac{2}{3}]$ with the U(1)-orbits in $f^{-1}(-\frac{2}{3})$ and $f^{-1}(\frac{2}{3})$ collapsed to points. However, the symplectic cut construction shows that it is a symplectic manifold and that the residual U(1)-action is Hamiltonian. Since this action is globally defined, its moment map $\hat{f}\colon \hat{M}_g \to [-\frac{2}{3},\frac{2}{3}]$ is perfect. It again has three critical submanifolds $\hat{S}_1, \hat{S}_2$, and $\hat{S}_3$. Of these, $\hat{S}_2 = S_2$ obviously. However, $\hat{S}_1$ and $\hat{S}_3$ are $S^2\times S^2$-bundles over M_{g-1}; this follows from Lemma 3.2.

Filter M_g by the open subsets $U_0 = f^{-1}(-\frac{3}{2},-\frac{3}{2})$, $U_1 = f^{-1}(-\frac{3}{2},-\frac{1}{2})$, $U_2 = f^{-1}(-\frac{3}{2},\frac{1}{2})$, and $U_3 = f^{-1}(-\frac{3}{2},\frac{3}{2})$, so that

$$\emptyset = U_0 \subset U_1 \subset U_2 \subset U_3 = M_g.$$

The machinery of Section 2 then produces a spectral sequence abutting to $H^*(M_g)$. We need to show that all the differentials are zero. Since there are only three nonzero terms in the cochain filtration, only the first two differentials can possibly be nonzero.

As shown in Section 2, the first differential is the direct sum of the maps $H^*(S_1)\to H^*(S_2)$ and $H^*(S_2)\to H^*(S_3)$ induced by the upper route in the diagram (2.1). We will consider these maps separately.

3.5. The map $H^*(S_1)\to H^*(S_2)$. As seen in 1.5, $H^*(S_1) = V \oplus \tau V$, where $\tau \in H^3$ is Poincaré dual to the locus where $B_g = I$, and $V = \pi^* H^*(M_{g-1})$. The map $H^*(S_1)\to H^*(S_2)$ accordingly splits into two maps $V \to H^*(S_2)$ and $\tau V \to H^*(S_2)$; we will show that each of these vanishes.

Let $U'_1 = f^{-1}(-\frac{2}{3}, -\frac{1}{2})$. Then $U'_1 \subset U_1$, but there is also a natural inclusion $U'_1 \subset \hat{U}_1$, where $\hat{U}_1 = \hat{f}^{-1}(-\frac{3}{2}, -\frac{1}{2})$. It then follows from Lemma 3.2 that the hexagon in the diagram below commutes. Here the maps of the second column are the Thom isomorphisms, which in this case are simply induced by retractions, and those of the first column are induced by the fiber bundle projections:

$$\begin{tikzcd}
 & H^*(S_1) \arrow[r] & H^*(U_1) \arrow[dr] & & & \\
H^*(M_{g-1}) \arrow[ur] \arrow[dr] & & & H^*(U'_1) \arrow[r] & H^*(U'_2, U'_1) \arrow[r] & H^*(S_2) \\
 & H^*(\hat{S}_1) \arrow[r] & H^*(\hat{U}_1) \arrow[ur] & & &
\end{tikzcd}$$

On the other hand, if $U'_2 = f^{-1}(-\frac{2}{3}, \frac{1}{2})$, then by excision $H^*(U'_2, U'_1) = H^*(U_2, U_1)$. The map $V \to H^*(S_2)$ in question is therefore the upper route in the diagram. However, the map $H^*(\hat{S}_1) \to H^*(S_2)$ induced by the lower route is a component of $\hat{d}_1$, the differential for $\hat{M}_g$, which vanishes since $\hat{f}$ is perfect. Hence the map $V \to H^*(S_2)$ must vanish.

For the second component, the map $\tau V \to H^*(S_2)$, this argument no longer works. Instead, we resort to a parity argument.

There is a symplectomorphism $\rho : M_g \to M_g$ induced by a half-twist of the gth handle of the surface X. Explicitly, it is given by $A_g \mapsto A_g B_g A_g^{-1} B_g^{-1} A_g^{-1}$ and $B_g \mapsto A_g B_g^{-1} A_g^{-1}$, with all other coordinates remaining fixed. This fixes f and hence preserves all of the sets discussed above. It acts as an involution on S_1, preserving the projection to M_{g-1} and the locus where $B_g = I$, but reversing orientation. It therefore takes τ to $-\tau$, so it acts as -1 on τV. On the other hand, it fixes S_2, so it acts trivially on $H^*(S_2)$.

A more subtle point is that ρ^* commutes with the Thom isomorphism $H^*(U_2, U_1) \to H^*(S_2)$. This follows from the Morse lemma provided that ρ preserves the orientation of the negative normal bundle. Of course, ρ will not preserve every Riemannian metric, so it may not preserve the corresponding negative normal bundle. But it does commute with the U(1)-action: this can be verified directly, or deduced from the fact that it preserves f and the symplectic form. If a suitable metric is chosen – for example, one equivalent near S_2 to a U(1)-invariant flat product metric on the normal bundle, via the equivariant symplectic tubular neighborhood theorem – then its negative normal bundle is the -1-weight space for the U(1)-action on $TM_g|_{S_2}$. This subbundle, and its natural orientation coming from the U(1)-action, are therefore preserved by ρ.

Any element of τV is therefore acted on as -1 in the first entry, and hence in the last entry, of the sequence of natural maps

$$H^*(S_1) \to H^*(U_1) \to H^*(U_2, U_1) \to H^*(S_2).$$

Its image in $H^*(S_2)$ must therefore be zero.

3.6. The map $H^*(S_2) \to H^*(S_3)$. This map is handled exactly like the previous one. The fact that $\hat{f}$ is perfect implies the vanishing of the component $H^*(S_2) \to \tau V$. Indeed, there is a diagram

$$\begin{tikzcd}
 & & & & H^*(U_3', U_2') \arrow[r] & H^*(S_3) \arrow[dr] & \\
H^*(S_2) \arrow[r] & H^*(U_2, U_1) \arrow[r] & H^*(U_2) \arrow[r] & H^*(U_2') \arrow[ur] \arrow[dr] & & & H^*(M_{g-1}) \\
 & & & & H^*(\hat{U}_3', \hat{U}_2') \arrow[r] & H^*(\hat{S}_3) \arrow[ur] &
\end{tikzcd}$$

where $U_3' = f^{-1}(-\frac{2}{3}, \frac{3}{2})$, $\hat{U}_3' = \hat{f}^{-1}(-\frac{2}{3}, \frac{3}{2})$, and the final column consists of push-forwards. It commutes because the map $H^*(U_2') \to H^*(S_3)$ can be interpreted as restriction to $f^{-1}(\frac{1}{3})$, which is a sphere bundle over S_3, followed by push-forward by projection, and similarly for $H^*(U_2') \to H^*(\hat{S}_3)$. But the push-forward $H^*(S_3) \to H^*(M_{g-1})$ is projection on the factor τV.

On the other hand, the parity argument implies the vanishing of the component $H^*(S_2) \to V$. Indeed, we have already seen that ρ^* commutes with all of the arrows between $H^*(S_2)$ and $H^*(U_3, U_2)$ in the above diagram. However, the Thom isomorphism $H^*(U_3, U_2) \to H^*(S_3)$ does not commute with ρ^*. Rather, the two composites differ by a sign. This is because ρ reverses the orientation of S_3, but not that of the open set U_3; after all, it is a symplectomorphism. It therefore reverses the orientation of the normal bundle to S_3, so it acts as -1 on the Thom class. Since any class in $H^*(S_2)$ is acted on trivially by ρ, its image in $H^*(S_3)$ must therefore be acted on as -1, so its component in V must vanish.

3.7. The map $H^*(S_1) \to H^*(S_3)$. At this point we have shown that the first differential vanishes. The second differential is therefore the map $H^*(S_1) \to H^*(S_3)$ given by taking $j = 1$ in the description of Section 2, and it suffices to show that this also vanishes. This time, since both $H^*(S_1)$ and $H^*(S_3)$ split as $V \oplus \tau V$, the differential splits into four components.

Of these, the component $V \to \tau V$ vanishes by the perfect argument. Simply note that, since $H^*(U_3, U_2) = H^*(U_3', U_2')$ by excision, a prime may be added to every set in the middle row of 2.1 without changing the differential. Then graft on the hexagonal diagrams of 3.5 and 3.6 and argue as before.

On the other hand, the component $V \to V$ vanishes by the parity argument. Indeed, suppose that a class in $H^*(S_1)$ belongs to the component V, so that it is invariant under ρ^*. Then the same is true of its image in $H^*(U_1)$. Since the differential $H^*(S_1) \to H^*(S_2)$ vanishes, this lifts to a class $u \in H^*(U_2)$. Now u may not be invariant under ρ^*, but the equivariance of the middle row of (2.1) implies that $\rho^* u = u + v$, where v is a class in the image of $H^*(U_2, U_1)$. Since the differential $H^*(S_2) \to H^*(S_3)$ vanishes, the image of v in $H^*(U_3, U_2)$ must be zero. Hence u maps to a class in $H^*(U_3, U_2)$ invariant under ρ^*. Its image in $H^*(S_3)$ is therefore anti-invariant as in 3.6, so its component in V vanishes.

Likewise, the component $\tau V \to \tau V$ vanishes by the same parity argument, with the roles of invariance and anti-invariance interchanged.

This leaves only one of the four components, namely the map $\tau V \to V$. Now τ is the restriction to S_1 of a global class $\tau \in H^3(M_g)$, namely the Poincaré dual of the locus where $B_g = I$. Furthermore, the differential d_2 is a module homomorphism over $H^*(M_g)$. Indeed, multiplication by a class such as τ clearly commutes with the natural maps shown as horizontal and vertical arrows in (2.1). It is therefore compatible with the lifting from $H^*(U_1)$ to $H^*(U_2)$: given a lifting $u \in H^*(U_2)$ of a class $w \in H^*(U_1)$, τu is a lifting of τw. And it is compatible with the Thom isomorphisms, since they are simply given by the cup product with the Thom class. Consequently, for any $v \in V$, $d_2(\tau v) = \tau d_2(v)$, but $d_2(v)$ has already been shown to vanish. This completes the proof of the main theorem.

4. The case of additional punctures

As promised in the introduction, the main theorem goes through for flat SU(2) connections on a surface X with more than one puncture. The proof is essentially the same. However, it is necessary to generalize the background material of Section 1 to the case of additional punctures.

So let $p_1, \dots, p_n$ be distinct points in X, let $t_1, \dots, t_n \in [0,1]$, let $\Gamma_1, \dots, \Gamma_n \subset \mathrm{SU}(2)$ be the conjugacy classes in SU(2) containing $\mathrm{diag}(e^{i\pi t_j}, e^{-i\pi t_j})$, and let $M_{g,n}$ denote the moduli space of flat SU(2) connections on X, or rather on $X \setminus \{p_1, \dots, p_n\}$, having holonomy around p_j in Γ_j. Of course $M_{g,n}$ depends on the choice of the t_j.

In analogy with Section 1, $M_{g,n}$ can be described as follows. Let $\mu_{g,n}: \mathrm{SU}(2)^{2g} \times \prod_{j=1}^n \Gamma_j \to \mathrm{SU}(2)$ be given by

$$\mu_{g,n}(A_1, B_1, A_2, B_2, \dots, A_g, B_g, C_1, \dots, C_n) = \left(\prod_{i=1}^{g} [A_i, B_i]\right)\left(\prod_{j=1}^{n} C_j\right).$$

Then $M_{g,n} = \mu_{g,n}^{-1}(I)/\mathrm{SU}(2)$.

Proposition 4.1. *For $J \subset \{1, \dots, n\}$, let $\kappa_J = \frac{1}{2}(\sum_{j \in J} t_j - \sum_{j \notin J} t_j)$. Then $\mu_{g,n}$ has $I \in \mathrm{SU}(2)$ as a critical value if and only if for some J, $\kappa_J \in \mathbb{Z}$.*

Proof. The derivative of $\mu_{g,n}$ at (A_i, B_i, C_j) is a map $D\mu_{g,n}: \mathfrak{su}(2)^{2g} \oplus \bigoplus_j T_{C_j}\Gamma_j \to \mathfrak{su}(2)$. A direct calculation shows that

$$\begin{aligned} D\mu_{g,n}(a_i, b_i, c_j) &= \mathrm{ad}\left(\prod_{j=1}^{n} C_j\right)^{-1} D\mu_{g,0}(a_i, b_i) + D\mu_{0,n}(c_j) \\ &= \sum_{i=1}^{g} \mathrm{ad}\left(\prod_{k>i} [A_k, B_k] \prod_{j=1}^{n} C_j\right)^{-1} \mathrm{ad}\, B_i A_i ((\mathrm{ad}\, B_i^{-1} - 1)a_i + (1 - \mathrm{ad}\, A_i^{-1}) b_i) \\ &\quad + \sum_{j=1}^{n} \mathrm{ad}\left(\prod_{\ell>j} C_\ell\right)^{-1} c_j. \end{aligned}$$

Suppose first that $g = 0$. Any $c_j \in T_{C_j}\Gamma_j$ is of the form $(1 - \mathrm{ad}\, C_j^{-1}) d_j$ where $d_j \in \mathfrak{su}(2)$. A further computation shows that

$$D\mu_{0,n}(c_j) = \sum_{j=1}^{n-1} \left(\mathrm{ad}\left(\prod_{k>j} C_k\right)^{-1} - 1\right)(d_{j+1} - d_j).$$

This fails to be surjective if and only if for all j, the image of $\mathrm{ad}(\prod_{k>j} C_k)^{-1} - 1$ is the same, which holds if and only if all $\prod_{k>j} C_k$ commute, which holds if and only if all C_j commute, since $\prod_{j=1}^n C_j = I$. The C_j may then be simultaneously diagonalized to $\mathrm{diag}(\mathrm{e}^{\pm i\pi t_j}, \mathrm{e}^{\mp i\pi t_j})$. Let J be the set of j for which the plus sign holds in the first factor. Then the stated condition on the t_j holds.

For $g > 0$, the presence of the term $(\mathrm{ad}\, B_i^{-1} - 1)a_i + (1 - \mathrm{ad}\, A_i^{-1})b_i$ in the formula above shows that if $D\mu_{g,n}$ is not surjective, then A_i and B_i must commute for each i. This brings us back to the case $g = 0$. $\square$

The case studied before was that of one puncture with $t_1 = 1$. Without loss of generality we may assume that, if we are not in this case, then each $t_j \in (0,1)$. Indeed, those j with $t_j = 0$ or 1 can be eliminated from the construction above without changing it, unless there are an odd number of j with $t_j = 1$. In the latter case, one can multiply some other C_j, say C_1, by $-I$, and change t_1 to $1 - t_1$.

Suppose that the t_j are so chosen that Proposition 4.1 is false, and I is a regular value. Then $M_{g,n}$ is smooth. A symplectic form on $M_{g,n}$ has been constructed by Guruprasad et al. [11]. Moreover, if $f: M_{g,n} \to [-1,1]$ and a U(1)-action on $f^{-1}(-1,1)$ are defined just as in Section 1, ignoring the extra C_j variables, then Audin [3] shows that this action is again symplectic with moment map $-i \arccos f$.

4.2. The critical submanifolds of f. The critical submanifolds of f are classified just as in 1.5. First, there are $f^{-1}(-1)$ and $f^{-1}(1)$, both of which are SU(2)-bundles over $M_{g-1,n}$ with vanishing Euler class. These are actually empty in certain cases when $g = 1$, but the main theorem will still hold even then.

The remaining critical points are again fixed points of the U(1)-action. It is straightforward to check that these are all conjugate to $(2g+n)$-tuples such that

$$A_g = \begin{pmatrix} \mathrm{e}^{-i\pi\kappa} & 0 \\ 0 & \mathrm{e}^{i\pi\kappa} \end{pmatrix}, \qquad B_g = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

and the remaining A_i, B_i, and C_j are all diagonal. For $i < g$, the A_i and B_i are free to move in the 1-parameter subgroup of diagonal matrices, but each C_j is in a fixed conjugacy class distinct from $\pm I$ and so must equal $\mathrm{diag}(\mathrm{e}^{\pm i\pi t_j}, \mathrm{e}^{\mp i\pi t_j})$. Let J be the set of those j for which the upper sign holds. The constraint imposed by μ then implies that $\kappa = \kappa_J$, as defined in Proposition 4.1. The critical submanifolds in $f^{-1}(-1,1)$ therefore consist of a disjoint union of 2^n tori of dimension $2g-2$.

Theorem 4.3. *The map f is a perfect Bott–Morse function on $M_{g,n}$.*

Proof. The proof is similar to that of the main theorem. We indicate only the points where it must be modified.

First of all, the three lemmas at the beginning of Section 3 go through unchanged, except that the formula in the proof of Lemma 3.1 gets replaced by

$$D\mu_{g,n} = D\mu_{g-1,n} + \mathrm{ad}\left(\prod_j C_j\right)^{-1} D\mu_{1,0}.$$

Suppose that $f^{-1}(\pm 1) \neq \emptyset$, since the result follows directly from Theorem 1.3 otherwise. Write the set of critical points as a disjoint union of submanifolds $S_1, \dots, S_N$ so that

$$-1 = f(S_1) < f(S_2) < \cdots < f(S_{N-1}) < f(S_N) = 1.$$

If there are no coincidences between the values of κ_J obtained in 4.2 for different J, then N will equal $2^n + 2$. Otherwise, some of the S_i will be disjoint unions of tori. The Morse indices of the components may vary, so the Thom class will be inhomogeneous, but this makes no difference.

Since there are N critical values, only the first $N-1$ differentials in the spectral sequence can possibly be nonzero. By induction suppose all differentials before d_k vanish. If $1 < i < k+i < n$, then the component $H^*(S_i) \to H^*(S_{k+i})$ of d_k vanishes because $-i \arccos f$ is a moment map there. The component $H^*(S_1) \to H^*(S_{k+1})$ is handled as in 3.5, except that in the parity argument, the ρ-equivariance of the long exact sequence of the relative cohomology of the pair (U_{i+1}, U_i) must be applied inductively as in 3.7 to show that the differential preserves the -1-weight space. Likewise the component $H^*(S_{N-k}) \to H^*(S_N)$ is handled as in 3.6. Finally, if $k = N-1$, then $d_k : H^*(S_1) \to H^*(S_N)$ is handled as in 3.7. □

This leads to a formula for the Betti numbers of $M_{g,n}$ as soon as the Morse indices of the critical tori are computed. At this point our approach meets the route taken by Hoyle [14]. From a lengthy trigonometric computation, he deduced the following formula.

Lemma 4.4 (Hoyle). *For any $J \subset \{1, \dots, n\}$, the index of the critical torus associated to J is $2g + 2n - 2|J| + 4[\kappa_J]$.*

Therefore, the Poincaré polynomials of $M_{g,n}$ satisfy the recursion

$$P_t(M_{g,n}) = P_t(S_1) + t^3 P_t(S^3) + (1+t)^{2g-2} \sum_J t^{2(g+n-|J|+2[\kappa_J])}$$

$$= (1+t^3)^2 P_t(M_{g-1,n}) + (t+t^2)^{2g-2} \sum_J t^{2(n+1-|J|+2[\kappa_J])}.$$

This immediately implies the following.

Corollary 4.5. *The Poincaré polynomial of $M_{g,n}$ satisfies*

$$P_t(M_{g,n}) = (1+t^3)^{2g} P_t(M_{0,n}) + \frac{(1+t^3)^{2g} - (t+t^2)^{2g}}{(1-t^2)(1-t^4)} \sum_J t^{2(n+1-|J|+2[\kappa_J])}.$$

From this, one can obtain an explicit expression for $P_t(M_{g,n})$ by substituting Hoyle's formula for $P_t(M_{0,n})$.

4.6. Remark. Assuming the Harder–Narasimhan formula, the recursion in the genus could also be proved by algebro-geometric methods. Here is a sketch of the argument. By a theorem of Mehta–Seshadri [21], $M_{g,n}$ may be regarded as a moduli space of vector bundles over X with parabolic structure at each p_j. If $t_n = 1$ but the remaining t_j are small, then $M_{g,n}$ is a $(\mathbb{CP}^1)^{n-1}$-bundle over M_g for $g > 0$, so the recursion holds in this case. For general t_j, choose a path in $[0,1]^n$

connecting $(t_1, \dots, t_n)$ to the previous special case in such a way that for any point on the path, $\kappa_J \in \mathbb{Z}$ for at most one $J \subset \{1, \dots, n\}$. Then the moduli spaces on either side of each such point are related by a blow-up and blow-down centered on a disjoint union of tori [4, 24]. It can be checked that this alters the Betti numbers in a way which preserves the recursion. Unlike the Morse approach, however, this algebraic approach gives no insight about why the recursion holds.

5. Some final remarks

Many other cheerful facts about $M_{g,n}$ can be deduced from the argument used to prove the main theorem. Let us mention two of them.

Proposition 5.1. *The integral cohomology of $M_{g,n}$ is torsion-free.*

Proof. When $g = 0$, then Hoyle [14] exhibits a Hamiltonian circle action on $M_{0,n}$ whose fixed points are isolated except for two copies of $M_{0,n-1}$. By induction we may suppose these are torsion-free, and Frankel [7, Corollary 1] then shows that $M_{0,n}$ is torsion-free.

Now suppose $g > 0$ and consider again the Bott–Morse function f. The spectral sequence of Section 2 works equally well with integral cohomology. Everything in the proof of the main theorem goes through, except that the differentials over $\mathbb{Z}$ might take nonzero values in the torsion part of $H^*(S_2, \mathbb{Z})$ or $H^*(S_3, \mathbb{Z})$. But S_2 is a torus, so it is certainly torsion-free, and S_1 and S_3 may be assumed torsion-free by induction on g. $\square$

Proposition 5.2. *Let $(\mathbb{Z}/2)^{2g}$ act on $M_{g,n}$ by*

$$(\delta_i, \varepsilon_i) \cdot (A_i, B_i, C_j) = ((-1)^{\delta_i} A_i, (-1)^{\varepsilon_i} B_i, C_j).$$

Then the induced action on $H^*(M_{g,n}, \mathbb{Q})$ is trivial.

Proof. The action of $(\mathbb{Z}/2)^{2g-2}$ on the first $2g-2$ factors preserves the U(1)-action, the map f and so on. The whole proof of the main theorem can therefore be $(\mathbb{Z}/2)^{2g-2}$-graded. But since $(\mathbb{Z}/2)^{2g-2}$ acts trivially on τ and on $H^*(S_2)$, by induction the whole grading is trivial. Hence $(\mathbb{Z}/2)^{2g-2}$ acts trivially on $H^*(M_{g,n})$.

The last two factors are really no harder. After all, the choice of an ordering on the handles of X was arbitrary. For example, the argument still works if f is replaced by $\frac{1}{2}\mathrm{tr}\, A_1$, and so on. $\square$

The value of this last result is to relate the cohomology of $M_{g,n}$ to that of the corresponding moduli space of flat U(2) connections. Let $\tilde{M}_{g,n}$ be the moduli space of flat U(2) connections on $X \backslash \{p_1, \dots, p_n\}$ with holonomy around p_j in the conjugacy class Γ_j. Just as $M_{g,n}$ was, $\tilde{M}_{g,n}$ may be described in terms of a map $\tilde{\mu}_{g,n}$.

Corollary 5.3. *As rings, $H^*(\tilde{M}_{g,n}, \mathbb{Q}) \cong H^*(\mathrm{U}(1)^{2g}, \mathbb{Q}) \otimes H^*(M_{g,n}, \mathbb{Q})$.*

Proof. There is a natural map $\mathrm{U}(1)^{2g} \times M_{g,n} \to \tilde{M}_{g,n}$ given by

$$(\kappa_i, \lambda_i), (A_i, B_i, C_j) \mapsto (\kappa_i A_i, \lambda_i B_i, C_j).$$

Indeed, $\tilde{M}_{g,n} = (\mathrm{U}(1)^{2g} \times M_{g,n})/(\mathbb{Z}/2)^{2g}$, where $(\mathbb{Z}/2)^{2g}$ acts diagonally on $\mathrm{U}(1)^{2g}$ and $M_{g,n}$ as above. The induced action on $H^*(\mathrm{U}(1)^{2g}, \mathbb{Q})$ is certainly trivial, and the induced action on $H^*(M_{g,n}, \mathbb{Q})$ is trivial by the lemma above. The proof is completed by the result of Grothendieck [10] that the rational cohomology ring of a quotient by a finite group is the invariant part of the rational cohomology. $\square$

Compare Newstead [23], Harder-Narasimhan [13] and Atiyah–Bott [1] in the case $n = 0$.

Acknowledgements

I am very grateful to Kenji Fukaya, Mark Hoyle, Michael Hutchings, Lisa Jeffrey, Tom Mrowka, and Jonathan Weitsman for helpful discussions. I especially thank Hoyle for providing me with a copy of his paper [14] in advance of publication.

References

[1] M.F. Atiyah, R. Bott, The Yang–Mills equations over Riemann surfaces, Philosophical Transactions of the Royal Society of London Series A 308 (1982) 523–615.
[2] M. Audin, The Topology of Torus Actions on Symplectic Manifolds, Progress in Mathematics, vol. 93, Birkhäuser, Basel, 1991.
[3] M. Audin, Lectures on gauge theory and integrable systems, in: J. Hurtubise et al., (Eds.), Gauge Theory and Symplectic Geometry, Kluwer, Dordrecht, 1997.
[4] H.U. Boden, Y. Hu, Variations of moduli of parabolic bundles, Mathematische Annalen 301 (1995) 539–559.
[5] R. Bott, An application of the Morse theory to the topology of Lie-groups, Bulletin of the Mathematical Society of France 84 (1956) 251–281.
[6] R. Bott, L.W. Tu, Differential Forms in Algebraic Topology, Graduate Texts in Mathematics, vol. 82, Springer, Berlin, 1982.
[7] T. Frankel, Fixed points on Kähler manifolds, Annals of Mathematics 70 (1959) 1–8.
[8] W. Goldman, The symplectic nature of fundamental groups of surfaces, Advances in Mathematics 54 (1984) 200–225.
[9] W. Goldman, Invariant functions on Lie groups and Hamiltonian flows of surface group representations, Inventiones Mathematicae 85 (1986) 263–302.
[10] A. Grothendieck, Sur quelques points d'algèbre homologique, Tôhoku Mathematical Journal 9 (1957) 119–221.
[11] K. Guruprasad, J. Huebschmann, L. Jeffrey, A. Weinstein, Group systems, groupoids, and moduli spaces of parabolic bundles, Duke Mathematical Journal 89 (1997) 377–412.
[12] G. Harder, Eine Bemerkung zu einer Arbeit von P. E. Newstead, Journal für die Reine und Angewandte Mathematik 242 (1970) 16–25.
[13] G. Harder, M.S. Narasimhan, On the cohomology groups of moduli spaces of vector bundles over curves, Mathematische Annalen 212 (1975) 215–248.
[14] M. Hoyle, Perfect Morse functions on the moduli space of parabolic bundles, preprint.
[15] J. Igusa, On a property of commutators in the unitary group, Memoirs of the College of Science of Kyoto University Series A 26 (1950) 45–49.
[16] L.C. Jeffrey, J. Weitsman, Toric structures on the moduli space of flat connections on a Riemann surface: volumes and the moment map, Advances in Mathematics 106 (1994) 151–168.
[17] L.C. Jeffrey, J. Weitsman, A perfect Morse function for the moduli space of flat connections, Compositio Mathematica 105 (1997) 147–151.

[18] Y. Karshon, An algebraic proof for the symplectic structure of moduli space, Proceedings of the American Mathematical Society 116 (1992) 591–605.
[19] F.C. Kirwan, Cohomology of Quotients in Algebraic and Symplectic Geometry, Mathematical Notes, vol. 31, Princeton, 1984.
[20] E. Lerman, Symplectic cuts, Mathematical Research Letters 2 (1995) 247–258.
[21] V.B. Mehta, C.S. Seshadri, Moduli of vector bundles on curves with parabolic structures, Mathematische Annalen 248 (1980) 205–239.
[22] M.S. Narasimhan, C.S. Seshadri, Stable and unitary vector bundles on a compact Riemann surface, Annals of Mathematics 82 (1965) 540–567.
[23] P.E. Newstead, Topological properties of some spaces of stable bundles, Topology 6 (1967) 241–262.
[24] M. Thaddeus, Geometric invariant theory and flips, Journal of the American Mathematical Society 9 (1996) 691–723.
[25] M. Thaddeus, An introduction to the topology of the moduli space of stable bundles on a Riemann surface, in: Geometry and Physics (Aarhus, 1995), Lecture Notes in Pure and Appl. Math., vol. 184, Marcel Dekker, New York, 1997.
[26] A. Weinstein, The symplectic structure on moduli space, in: H. Hofer et al. (Eds.), The Floer Memorial Volume, Progress in Mathematics, vol. 133, Birkhaüser, Basel, 1995.